

1.	Course Title	Mining Massive Data Sets	
2.	Code	F18L3W154	
3.	Study program	Software engineering and information systems	
4.	Study Program Organizer	Faculty of Computer Science and Engineering	
5.	Degree (first, second, third cycle)	first cycle	
6.	Academic year / semester 4 / winter / optional	7. ECTS credits	6
8.	Teacher	associate professor Slobodan Kalajdzhiski, associate professor Gjorgji Madzharov, assistant professor Eftim Zdravevski	
9.	Course enrollment prerequisites	Паралелно и дистрибуирано процесирање	
10.	Course program goals (competencies): The course will introduce the students data mining and machine learning algorithms for analyzing massive amounts of data. The emphasis will be on the distributed platforms and Map Reduce as a tool for creating parallel algorithms that can process large amounts of data.		
11.	Course program content: Introduction and MapReduce. Association Rules: Frequent itemsets and Association rules, Near Neighbor Search in High Dimensional Data, Locality Sensitive Hashing (LSH), Dimensionality reduction: SVD and CUR, Recommendation Systems, Clustering, Random Walks with Restarts, Large scale supervised machine learning: k-nearest neighbor, Perceptron, Classification and regression trees, Mining data streams, Web Advertising.		
12.	Learning methods: Lectures using presentations, interactive lectures, exercises (using equipment and software packages), teamwork, case studies, invited guest lecturers, independent preparation and defense of a project assignment and seminar work.		
13.	Total available time	180	
14.	Distribution of the available time	30 + 30 + 19 + 15 + 75 = 169 hours	
15.	Teaching activity forms	15.1. Lectures – theoretical teaching	30 hours
		15.2. Exercises (laboratory, auditory), seminar papers, teamwork	30 hours

16.	Other activity forms		16.1.	Project Tasks	19 hours	
			16.2.	Independent Learning Tasks	15 hours	
			16.3.	Home learning	75 hours	
17.	Assessment methodology					
	17.1.	Tests			0 points	
	17.2.	Seminar paper/project (presentation: written and oral)			40 points	
	17.3.	Activity and learning			0 points	
	17.4.	Final exam			60 points	
18.	Assessment criteria (points/grade)		up to 50 points		5 (five) (F)	
			51 to 60 points		6 (six) (E)	
			61 to 70 points		7 (seven) (D)	
			71 to 80 points		8 (eight) (C)	
			81 to 90 points		9 (nine) (B)	
			91 to 100 points		10 (ten) (A)	
19.	Course completion and final exam requirements		Realized activities 15.1 and 15.2			
20.	Teaching Language		Macedonian and English			
21.	Teaching quality evaluation method		Internal evaluation mechanisms and questionnaires			
22.	Course Material					
	22.1.	Mandatory course material				
		No	Author	Title	Publisher	Year
		1	Anand Rajaraman and Jeffrey Ullman	Mining of Massive Datasets	Cambridge University Press	2014
		2	Avrim Blum, John Hopcroft, and Ravindran Kannan	Foundations of Data Science	Draft version	2017
		3	Jiawei Han, Micheline Kamber, and Jian Pei	Data Mining: Concepts and Techniques, Third Edition	Morgan Kaufmann	2011
	22.2.	Additional course material				
		No.	Author	Title	Publisher	Year

